UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/670,681	09/25/2003	Eric D. Brill	MS305080.1/MSFTP475US	7563

27195   7590   12/31/2007
AMIN. TUROCY & CALVIN, LLP
24TH FLOOR, NATIONAL CITY CENTER
1900 EAST NINTH STREET
CLEVELAND, OH 44114

EXAMINER
JEAN GILLES, JUDE

ART UNIT	PAPER NUMBER
2143	

NOTIFICATION DATE	DELIVERY MODE
12/31/2007	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

docket1@thepatentattorneys.com
hholmes@thepatentattorneys.com
osteuball@thepatentattorneys.com

PTOL-90A (Rev. 04/07)

	Application No.	Applicant(s)
Office Action Summary	10/670,681	BRILL ET AL.
	Examiner	Art Unit	
	Jude J. Jean-Gilles	2143	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _28 September 2007_.

2a)☒ This action is **FINAL**.   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112, 114-116_ is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103, 105-112 and 114-116_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on _25 September 2003_ is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

This office action is responsive to the Reply filed on 09/28/2007.

Response to Amendment/Arguments

1.   Claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103,105-112,

114-116 are currently pending in the subject application and are presently under

consideration. Claims 1, 26, 37, 57, 60, 61, 65, 76, 78-82, 84-92, 102, 103, 105,108

and 114 have been amended, and . Claims 8, 22, 45, 48,, 66, 77, 93, 94, 101,104 and

113 have been cancelled. Pending claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92,

95-100, 102, 103,105-112, 114-116 represent "SYSTEMS AND METHODS FOR

CLIENT-BASED WEB CRAWLING".

Applicant's arguments with respect to claims 1-116 have been carefully

considered, but are not deemed fully persuasive. Applicant's arguments are deemed

moot in view of the new ground of rejection as explained here below, necessitated by

Applicants' substantial amendment to the claims.

Examiner notes that applicant has failed in presenting claims and drawings that

delineate the contours of this invention as compared to the cited prior art. Applicant has

failed to clearly point out patentable novelty in view of the state of the art disclosed by

the references cited that would overcome the 103(a) rejections applied against the

claims, the rejection is therefore sustained.

Claim Rejections - 35 USC § 103

2.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

3.   **Claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103,105-112**

and 114-116 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bailey et

al (Bailey), Pub. No. 20060167864 A1 in view of Albion et al. (hereinafter Albion) US.

Pub. No. 20040240388 A1.

Regarding claim 1: Bailey discloses the invention substantially as claimed. Bailey

teaches a data analysis system (fig. 1), comprising:

a first component <u>associated with a server of the data analysis system</u> that

facilitates generation of a first data set related to web page information obtained via a

communication system (fig. 1, item 120; par. 0029); and

a second component that coordinates a data set relating to web page information

from at least one distributed resource which interacts with the communication system;

the second data set is utilized to refine the first data set (see abstract, fig. 1; *note that*

the web crowler (160) generates the data set through the Internet; 0037-0040, 0052).

However Bailey does not specifically discloses a system<u>, wherein refining the first data</u>

<u>set comprises adding unknown information to the first data set, when new information is</u>

<u>received from the distributed source via the second data set or updating existing</u>

<u>information in the first data set when changes have occurred in the contents of the web</u>

<u>page information as indicated by the second data set.</u> Nonetheless, this feature is well

known and would have been an obvious modification to the system shown by Bailey as

evidenced by Albion.

In an analogous art, Albion shows a plurality of clients capable sending updates

information to a crawler, this request information is the second data which updates the

timer list with new data (see Albion, par. 0020-0022). In an attempt to update the web

crawler with new information from the distributed source, change information is passed

from the clients to the crawler with update timer list contents.

Given this feature, a person of ordinary skill in the art would have been readily

recognized the desirability and advantages of modifying the system shown by Bailey to

employ the features shown by Albion in order to facilitate the dynamic content updates

(see Albion, par. 0001, 0011). In referring to 2, Albion shows a crawler receiving timer

information from a client and updates its timer list in memory with the new information.

By this rationale, claim 1 is rejected.

Regarding claims 1-7, 9-21, 23-44, 46, 47, 49-65, 67-76, 78-92, 95-100, 102, 103,105-112

and 114-116, the combination Bailey-Albion teaches:

2. The system of claim 1, the first component comprising an internet web crawler (see

Bailey; 120).

3. The system of claim 1, the first component comprising an intranet web crawler (see Bailey; 120; *the crawler is usable equally in the Internet, as well as an Intranet*).

4. The system of claim 1, the second component further utilized to optimize reception of data from the distributed resources (see Bailey; 164).

5. The system of claim 1, the second component provides a scheduling function to control reception of the second data set from the at least one distributed resource (see Bailey; 147).

6. The system of claim 1, the second component utilized to facilitate communication traffic reduction via the communication system by employing a proper set of weak indicator functions representative of the first data set (see Bailey; 162).

7. The system of claim 6, the second component further utilized to randomly select and transmit a weak indicator function selected from the proper set of weak indicator functions to at least one of the distributed resources (see Bailey; 160, 162, 164).

9. The system of claim 1, the second component further utilized to generate status information about data related to the first data set; the status information transmitted to

at least one distributed resource (see Bailey; fig. 5; 0070).

10. The system of claim 9, the status information comprising, at least in part, a

freshness flag to indicate freshness of information related to the first data set (see

Bailey; fig. 5; 0070).

11. The system of claim 9, the status information comprising, at least in part, a hash of

contents of information related to the first data set (see Bailey; fig. 5; 0070, 0076).

12. The system of claim 9, the status information comprising, at least in part, a copy of

information of the first data set (see Bailey; fig. 5; 0070, 0076).

13. The system of claim 1, the communication system comprising an internet (see

Bailey; 110, 120, 130).

14. The system of claim 1, the communication system comprising a world wide web

(see Bailey; 110, 120, 130).

15. The system of claim 1, the communication system comprising an intranet (see

Bailey; 110, 120, 130).

16. The system of claim 15, the intranet comprising a local area network . (see Bailey;

130).

17. The system of claim 15, the intranet comprising a wide area network (see Bailey;

110, 120, 130).

18. The system of claim 1, the distributed resources comprising clients of a server (see

Bailey; 110, 120, 130).

19. The system of claim 1, the distributed resources comprising trusted entities

interactive with the communication system and the second component (see Bailey; fig.

2, 5,.

20. The system of claim 1, the first data set comprising internet web page data (see

Bailey; 0043, 0070, 0087; fig. 1 & 2).

21. The system of claim 1, the first data set comprising intranet web page data (see

Bailey; 0043, 0070, 0087; fig. 1 & 2).

23. The system of claim 1, the second data set comprising, at least in part, a hash of

contents of at least one web page (see Bailey; 0040, 0070, 0087; fig. 1, 2, & 5).

24. The system of claim 1, the second data set comprising, at least in part, a Uniform Resource Locator (URL) of at least one web page (see Bailey; 0040, 0070, 0087; fig. 1, 2 & 5).

25. The system of claim 1, the second data set comprising, at least in part, a time stamp relating to an acquisition time for information about at least one web page (see Bailey; 0043, 0070, 0087; fig. 1 & 2).

26. The system of claim 1, the second data set comprising, at least in part, a delta indication of the changes to contents of at least one web page (see Bailey; 0043, 0070, 0087; fig. 1 & 2).

27. The system of claim 26, the delta indication including, at least in part, a hash of previous contents of a web page and a hash of recent contents of the web page (see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

28. The system of claim 1, the second data set comprising, at least in part, a status indication of changes to contents of at least one web page(see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

29. The system of claim 28, the status indication including, at least in part, a percentage relating to an amount of change of contents of a web page (see Bailey; 0028, 0043,

0070, 0087; 0076, fig. 1 & 2).

30. The system of claim 28, the status indication including, at least in part, a

significance indicator to signify importance of changes in contents of a web page (see

Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

31. The system of claim 1, the second data set comprising internet web page data (see

Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

32. The system of claim 1, the second data set comprising intranet web page data (see

Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

33. The system of claim 1, the second data set comprising data compiled utilizing at

least one weak indicator function randomly selected from a set of weak indicator

functions; the set of weak indicator functions representative of the first data set (see

Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

34. The system of claim 1, further comprising a search component to accept at least

one search query and generate at least one search reply having at least a portion of the

first data set represented by information embedded in the search reply (see Bailey;

0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

35. The system of claim 1, further comprising a web page server component to construct web pages having at least a portion of the first data set represented by information embedded in at least one link found on at least one constructed web page (see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

36. The system of claim 1 further comprising a storage component to store the first data set (see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

37. A method for facilitating data analysis, comprising:

generating a first data set relating to a second data set obtained from web pages interactive with a server of a communication system (see Bailey; see abstract; fig. 1; (see Bailey; 0037-0040, 0052);

receiving a third data set from at least one distributed resource that is interactive with the communication system; the third data set comprising web page related information generated by the distributed resource; and refining the second data set to reflect information obtained from the third data set (see Bailey; 0084-0088);

adding unknown information to the second data set when new information is received from the distributed source via the third data set;

updating existing information in the second data set when changes have occurred as indicated by the third data set; and

passing status information to the distributed resource through one or more indicators

after information from the third data set has been analyzed (see Albion, par. 0020-

0022).

38. The method of claim 37, the first data set comprising a representation of the second

data set (see Bailey; see abstract; fig. 1; 0037-0040, 0052).

39. The method of claim 38, the representation of the second data set comprising, at

least in part, a hash of contents of at least one web page contained in the second data

set (see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

40. The method of claim 38, the representation of the second data set comprising, at

least in part, a status indication of at least one web page contained in the second data

set (see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

41. The method of claim 40, the status indication comprising a freshness flag to indicate

if the web page information is current (see Bailey; fig. 5; 0070).

42. The method of claim 37, the first data set comprising a copy of the second data set

(see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

43. The method of claim 37, the second data set comprising web page information

compiled by a web crawler (see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

44. The method of claim 37, the third data set comprising web page information based upon client accessed web page information on the communication system (see Bailey; 0028, 0043, 0070, 0087; 0076, fig. 1 & 2).

46. The method of claim 37, the communication system comprising an internet (see Bailey; fig. 1).

47. The method of claim 37, the communication system comprising an intranet (see Bailey; fig. 1).

49. The method of claim 37, further including: transmitting the first data set to at least one distributed resource that is interactive with the communication system making the first data set available to be utilized by the distributed resource to generate the third data set (see Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

50. The method of claim 38, further including: generating a set of weak indicator functions to represent the second data set; and selecting random weak indicator functions from the set of weak indicator functions to transmit to the distributed resources as the first data set (see Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

51. The method of claim 50, the set of weak indicator functions comprising a proper set of weak indicator functions such that a non-zero probability exists that a randomly selected weak indicator function can identify a new web page (see Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

52. The method of claim 50, generating a set of weak indicator functions comprising: providing a dictionary representative of the second data set; partitioning randomly the dictionary into non-overlapping subdictionaries; and creating a function where $I(x)=1$ if and only if at least one subdictionary's weak indicator function is equal to one (see Bailey; 0076-0080).

53. The method of claim 37, further including: comparing the third data set to the second data set to reveal spoof data included in the second data set (see Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

54. The method of claim 37, further including: optimizing reception of at least one third data set through scheduling of the distributed resources (see Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

55. The method of claim 37, further including: receiving a web page search query from at least one distributed resource; generating a web search results page in response to

the web page search query from the distributed resource; embedding portions of the

first data set in links found on the web search results page; and transmitting the web

search results page as a representation of at least a portion of the second data set to

the distributed resource (see Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076,

fig. 1 & 2).

56. The method of claim 37, further including: constructing a web page utilizing at least

a portion of the first data set to embed information about links found in the web page;

and transmitting the web page to disseminate the first data set to at least one distributed

resource (see Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

57. A data analysis system, comprising:

means for generating at least one first data set from a <u>server of</u> communication system;

means for receiving and coordinating at least one second data set from at least one

~~distributed resource~~ client which interacts with the <u>server of the</u> communication system;

and means for refining the first data set utilizing at least one second data set (see

Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2);

wherein refining the first data set comprises the at least one of adding unknown

information to the first data set when new information is received from the client via the

second data set and updating existing information in the first data set when changes

have occurred in the web pare as indicated by the second data set (see Albion, par.

0020-0022).

61. A data analysis system, comprising: a first component <u>associated with at least one</u>

<u>client of a distributed web crawling system</u>

that generates web page information from at least one visited web site for utilization in

[a] <u>the</u> distributed web crawling system; ~~the web page information transmitted by the~~

~~first component to a second component~~; <u>and</u> (see Bailey; 0028, 0084-0088; 0037-0043,

0070, 0087; 0076, fig. 1 & 2)

a second component <u>associated with a server that receives the web page information</u>

<u>transmitted by the first component</u> *via* a communication system, <u>wherein the first</u>

<u>component receives a set of data from the second component to utilize in the</u>

<u>generation of the web page information comprising at least comparison data based on</u>

<u>the visited web page and the received set of data </u>(see Albion, par. 0020-0022).

92. A method for facilitating data analysis, comprising: compiling a first data set derived

from accessing web pages via a <u>client of a</u> communication system; ~~and~~ transmitting,

selectively, the first data set to an entity <u>comprising at least a server</u> of a distributed

crawling system that is interactive with the communication system (see Bailey; 0028,

0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2).

 <u>receiving a representation of a second data set compiled by the server of the web</u>

<u>crawler;</u>

<u>the second data set relating to at least one web page from the communication system;</u>

and utilizing the second data set to control which web pages to visit to compile the first

data set (see Albion, par. 0020-0022).

114. 114. (Currently Amended) A computer readable medium having stored thereon

computer executable components comprising:

a first component associated with a server of the data analysis system that

facilitates generation of a first data set related to web page information obtained *via a*

communication system; and a second component that coordinates a second data set

relating to web page information from at least one distributed resource associated with

at least a client of the server which interacts with the communication system; (see

Bailey; 0028, 0084-0088; 0037-0043, 0070, 0087; 0076, fig. 1 & 2)

the second data set is utilized to refine the first data set, wherein refining the first data

set comprises adding unknown information to the first data set when new information is

received from the distributed source via the second data set and updating existing

information in the first data set when changes have occurred in the contents of the web

page information as indicated by the second data set(see Albion, par. 0020-0022).

Claims 58-60, 62-65, 67-76, 78-92, 95-1100, 102, 103, 105-112, and 115-116 are

similar to other claims addressed above (see rejection of claims 2-56 above).

Conclusion

4.   Applicant's amendment necessitated the new ground(s) of rejection presented in

this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP

§ 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37

CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE

MONTHS from the mailing date of this action. In the event a first reply is filed within

TWO MONTHS of the mailing date of this final action and the advisory action is not

mailed until after the end of the THREE-MONTH shortened statutory period, then the

shortened statutory period will expire on the date the advisory action is mailed, and any

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later

than SIX MONTHS from the date of this final action.

6.   Any inquiry concerning this communication or earlier communications from

examiner should be directed to Jude Jean-Gilles whose telephone number is (571) 272-

3914. The examiner can normally be reached on Monday-Thursday and every other

Friday from 8:00 AM to 5:30 PM.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Nathan Flynn, can be reached on (571) 272-1915. The fax phone number

for the organization where this application or proceeding is assigned is (571) 273-3201.

Any inquiry of a general nature or relating to the status of this application or

proceeding should be directed to the receptionist whose telephone number is (703) 305-

0800.

SUPERVISORY PATENT EXAMINER
NATHAN FLYNN

Jude Jean-Gilles

Patent Examiner

Art Unit 2143

December 25, 2007